DOCUMENT RESUME

ED 335 866 FL 018 955

AUTHOR Wongsothorn, Achara
TITLE Steps in the Construction and Validation of Domain-Referenced Tests of English Use.
PUB DATE 90
NOTE 46p.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Standards; Communicative Competence (Languages); *English (Second Language); Foreign Countries; Higher Education; *Language Proficiency; *Language Skills; *Language Tests; Listening Skills; Reading Skills; Secondary Education; Speech Skills; Standardized Tests; *Test Construction; Test Validity; Writing Skills
IDENTIFIERS *Domain Referenced Tests; *Thailand

ABSTRACT

A project in Thailand to develop domain-referenced English language tests for use with Thai secondary and college students is described. The first phase consisted of developing standardized tests for listening, speaking, reading, and writing based on existing syllabuses, determining the common levels of English proficiency in those domains, and analyzing the communicative competence factors in English use of the Thai students. Subjects were 697 lower secondary, 525 upper secondary, and 493 university students. The second phase involved examining the consistency of domain specifications and setting domain-content standard scores for the tests. The process of test construction and validation are outlined, and results for each phase are reported. Recommendations for further research in this area include study of the contribution of individual language skills to the development of overall language proficiency, with the objective of finding effective remedial methods, and similar test and standard development for all levels of Thai education. Recommendations for practical application in English teaching and learning include use of a communicative context for language instruction and early emphasis on oral/aural skills gradually leading to later emphasis on reading and writing. A 55-item bibliography and test objectives, specifications, and analyses are appended. (MSE)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED335866



"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Wongsothorn

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☆ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

# STEPS IN THE CONSTRUCTION AND VALIDATION OF DOMAIN-REFERENCED TESTS OF ENGLISH USE

*BY*

*ACHARA WONGSOTHORN*

*CHULALONGKORN UNIVERSITY LANGUAGE INSTITUTE*

FL 018955

BEST COPY AVAILABLE

## *INTRODUCTION TO DOMAIN-REFERENCED TESTING*

Domain-referenced testing dates back to 1966 when Wells Hively of the University of Minnesota started the MINNEMAST Project aiming at the development of domain-referenced tests for science and mathematics. Publications on this subject coming out in 1974 and 1975 were influential especially in Minnesota to the widespread use of this new trend in testing. Domain-referenced testing expanded on the concept of criterion-referenced testing which measures students' mastery of the set objectives by building items on the incorporated domain description, content limits, criterion of distractor description, criterion of distractor domains, format and direction, and specimen items. Sension and Rabehl (1974) delineated the steps of domain-referenced test construction as consisting of:

1. developing the domain on the results of the analyses of contents/skills, teaching and learning procedures, objectives and learner-centered goals which are clear and easily understood, and

2. developing domain-referenced test items by sampling of testees and content domains, and setting suitable test techniques relevant to the teaching-learning situations.

The results of the tests are to be analyzed and reported to all parties concerned, i.e., teachers, students,

parents,etc. The report of the student's profile will, in turn, serve the purposes of remedial teaching and both short-term and long-term learning improvement. Information yielded by domain-referenced tests is greater than that yielded by criterion

2

referenced tests because it is directly linked to the domains of teaching and learning, thus, readily lending itself to remedial teaching and learning improvement. In other words, domain-referenced testing is a system developed to measure learning mastery which goes beyond criterion-referenced testing by not merely focusing on the criteria of achievement set by the behavioral objectives, but pinpointing also on the content /skill and the context of evaluation and the environment of learning and acquisition as a domain incorporates the areas covered by the objectives, the content of learning, the skills or sub-skill levels, for instance speaking skill at the criticism level, reading skill at the comprehension and interpretative levels. Student learning/testing activities, such as taking notes, listening to lectures, etc., and the learning/testing media, such as video--audio tapes, newspapers, OHP,etc. are included in the domain-description. A domain must have a definite area with a clear borderline separating it from other domains.

The amplified objectives on which a domain-referenced test specification is based are expanded from behavioral objectives which have been criticized as not being objective nor complex enough for measuring human learning. Specificity

is one of the necessary dimensions of amplified objectives. Roid and Haladyna (1982: 297) mentioned the effort to improve behavioral objectives by Popham of the University of California IOX (Item-Objective Exchange). Popham's amplified objectives consists of 5 parts as follows:

1. Description of the traits to be measured or the main objectives of testing,

3

2. Specific objectives which have been transformed from the main objectives,

3. Descriptions of item forms of both the stimuli and the responses; in other words, behaviors/performance to be tested, types and techniques of testing, the test items, format and techniques of testees's responses are specified,

4. Condition and criteria of correct responses,

5. Specimen test items.

Thorndike (1982: 2) described the process of developing domain-referenced test items as consisting of 5 steps as follows:

1. Specify and limit the domain,

2. Specify the test techniques or the format of the stimuli,

3. Specify the technique or mode of responses,

4. Specify the amount of items unbiasedly sampled from the universe of content,

5. Specify the criteria of mastery of the domain.

The descriptions of domain-referenced testing and

amplified objectives above confirm their behavioral roots. With the advent of individualized, learner-centered approach of the early 1970's, the domain-referenced item-form shell developed by Hively, et. al.(Alkin, 1973: 25) incorporated the transformation rule by which items can be substituted by paralleled ones, and the instructions to be given to individual students whose behaviors are to be observed and recorded. Lists of stimuli--verbal, pictorial, realia, etc.--to be



presented to the testees form a component of an item shell.

Purves, et. al.(1984) specified the system of giving instructions for essay writing in their international study on domain-referenced essay writing to include 15 dimensions, i.e., instructions, stimuli, cognitive demand, goal, role, audience, content, genre, style, advanced preparation of information, length, writing pattern, time, drafting and criterion. Purves' domain specification has proven useful for essay writing as it reminds the students of the details and complexities of the writing tasks as well as serving as a tool for evaluating an essay.

***Communicative Approach to Domain-referenced Testing***

Testing communicative language competence is based on socio-cultural and pragmatic theories of language learning. Language contents, encoding skills and subskills--unitary or integrative--are meshed together in the dynamics of communication with the following communicative parameters: function, modality, channel of communication, status, role, style, topic and communicative situations. An item shell of a

communicative domain-referenced test may include discrete-point items measuring linguistic competence, macro/global items measuring sociolinguistic competence, single/unitary and integrated-skill items measuring strategic, pragmatic and discourse competence. Testing methods may be either or both selection (multiple choice, true/false, matching, sequencing, etc.) and supply (cloze, completion, short-answer, letter-writing, essay, etc.).

Using B. J. Carroll's (1980) parameters for communicative test construction in developing a domain-referenced test the following parameters are included in the



test blueprint:

1. Context: time and location of language interaction, e.g., in the classroom, at the train-station, in the restaurant, etc.

2. Content Key: main idea or topic of language interaction, e.g., language in bargaining, in business transaction, in signing a treaty, in lecturing about danger of neuclear wars, in advertising, etc.

3. Function and Discourse: language for work or personal social transaction, elaborated/restricted code, referential/directive/poetic/phatic function or metalanguage, descriptive/persuasive/explanatory discourse.

4. Modality: language interactions via the oral mode (listening, speaking) or the visual mode (reading, writing).

5. Genre: methods of employing language skills, e.g., writing a personal letter, writing an essay, taking notes,

writing a telegram, talking on the telephone, reading a newspaper, etc.

6. Role: interaction between language transmitters and receivers, e.g., teacher-student, father-son, friend-friend, younger brother-elder sister, etc.

7. Status: socio-economic statuses or professional areas governing the style and register of language use, e.g., ecclesiastical language, court language, vulgar language, language for specific purposes, such as English for airline hostesses and stewards and stewardesses, etc.

8. Pre-supposition: anticipation of language transmissions, e.g., expectancy of the incoming message through



some media or a face-to-face interaction; familiarity with the language user or transmitter will facilitate communication, for example reading a friend's smeared letter with some words or phrases missing, or listening to one's mother in a noisy market place. Here communication is not hampered on account of the receivers correct anticipation of the incoming message.

9. Mood or Attitudes: the mood and attitudes of language users can determine the content and style of language use, e.g., humorous, rude, sincere, sarcastic, negative, positive, etc.

10. Formality: the level of language use, e.g., formal, colloquial, slang, academic, standard, sub-standard, etc.

The aforementioned framework for domain-referenced

testing within the communicative language teaching/learning approach is congruent with the English language teaching /learning practice currently endorsed in various sectors of the Thai educational systems where English is viewed as the language of international communication and where meanings of the message are the heart of the matter.

***THE RESEARCH***

The research on domain-referenced tests of English use was conducted on a grant given by the Thai government and was administratively supported by the Chulalongkorn University Language Institute and the Testing Bureau of the Ministry of Education. The research has two phases. Phase I Project was entitled Reearch for the Development of



English Domain-Referenced Tests for Use with Thai Students in Various Levels. The objectives of Phase I Project were to:

1. Develop standardized domain-referenced test-items from the determined domains of English use via the sound modality, i.e., listening and speaking and the graphic modality, i.e., reading and writing, as specified in the English syllabuses for the lower secondary, upper secondary and tertiary levels of education,

2. Determine the common levels of English proficiency in the specified domains of students in the three levels of education as stated in objective # 1, and to

3. Study, analyze and determine the factors of communicative competence in using English of Thai students in the lower secondary and tertiary levels of education.

The multi-stage stratified random sampling method was employed to obtain the sample groups of 697 lower secondary students, 525 upper secondary, and 493 Chulalongkorn University students.

The Rasch Model and the sensitivity and facility indices were used for test item analyses and revision. Domain-referenced methods of test analyses were used to establish the tests' coefficients of reliability and validity. F-tests and t-tests were used to establish test criterion-related validity.

Parametric statistics used to discover students' levels of English proficiency in the specified domains of English use were arithmetic mean, standard deviation, coefficient of variation(CV), kurtosis and skewness. Students'



proficiency levels were rated on a 5-band scales from very weak requiring extensive remedial work, to very good not requiring any remedial work.

To study the factors of communicative competence, Exploratory Principal Component Factor Analysis with Varimax Rotation and Pearson Product Moment Correlations were employed.

The research project in Phase II entitled Examination of the Consistency of Domain Specifications for English Language Test in the Lower and Upper Secondary Levels of Education aimed at:

1. Examining the consistency of domain specifications of the domain-referenced tests of English use, and

2. Setting the domain-content standard scores for the domain-referenced tests of English use constructed for the research. (Details in Appendix IV)

The sample groups consisted of 563 lower secondary and 469 upper secondary students selected by the stratified random sampling method.

Test reliability indices were established using the Cronbach Alpha Mehod. The Rasch Model was employed to examine the test items' difficulty indices and their congruence with the ICC or the item characteristic curve. To prove the consistency of domain specifications used to develop three forms of domain-refenced tests for each level totalling nine tests, the F-test method was used on the aritmetic means of item difficulty of each form. With the F ratio of 0.00 and the probability of 0.99, the consistency of the domain specifications for the three parallel forms--Form A, Form B, and Form C was confirmed.

9

The percentile method was used to set the test domain-content standard scores. (See Appendix IV)

The research in both of its phases is a language testing research in its full sense by being:

1. a research to develop three sets of domain-referenced tests of English use, three forms for each set, from the English curricula of the lower and upper levels of

secondary of education and from the English syllabuses of the Chulalongkorn University Language Institute for students in first, second and third years at Chulalongkorn University; the test items were processed through the various steps of test analyses for the purpose of test standardization and validation,

2. a research to find out about the nature and specific characteristics of communicative proficiency in the English language by the use of the constructed and standardized domain-referenced tests of English use,

3. a research using the domain-referenced tests of English use as a tool for studying the levels of English ability of Thai students in three levels of education.

***LANGUAGE TESTING RESEARCH IN THE FIRST SENSE***

To develop, standardize, and validate the domain-referenced tests of English use for the three levels of education--lower secondary, upper secondary, and tertiary--the following steps were taken:

***Step One***

From the learning objectives specified in the English



syllabuses amplified objectives were developed.

(See Appendix I for a specimen set of amplified objectives)

***Step Two***

From the amplified objectives domain-referenced test tables of specification were developed to sample the English learning domains set in the syllabuses and encapsulated in the

amplified objectives.

(See Appendix II for a specimen domain-referenced test table of specification)

***Step Three***

The multitrait-multimethod assessment scheme was used in test development. From the domain-referenced test tables of specification the multimethod approach to test construction consisting of both selection-type items, ie., multiple choice, matching, and item sequencing, and supply-type items, ie., completion of statements, diagrams and charts, short-answers, partial dictation, full dictation, summarization, paragraph writing and essays. The multitrait assessment consisted of measuring both language components, ie., grammar, vocabulary, and phonology which are discrete-point items, and language skills: unitary skills of listening, speaking, reading and writing and integrative skills of listening-writing, listening-reading-writing, and reading-writing. All the items were maximal performance tests except for the items measuring student speaking skill which employed a self-assessment rating scheme.

For the lower secondary level 225 items, 75 items per form, were constructed. For the upper secondary level also 225 items, 75 items per form, were developed. For the tertiary

11

level within the context of Chulalongkorn University, 210 items, 70 items per form, were constructed. The tests, therefore, consisted of 3 parallel forms for each level of education. Each form was based on the domain-referenced test table of

specification developed from the amplified objectives for each level.

*Step Four*

For test standardization the following methods of test analysis were performed on the students' obtained raw scores:

1. The classical model of item analysis, ie., Chung Teh Fan (1952) 27% upper-lower method was used to find out item difficulty and discrimination indices which were used as a basis for item revision. The Whitney & Sabers (1970, 1971) method was used on student scores from the multiple-point items, ie., essays, short-answers, partial dictation, full dictation, and summarization for the same purpose.

2. The second pre-tests were conducted using the revised test forms on the sampled subjects of the first pre-tests consisting of 42 lower-secondary, 41 upper-secondary, and 47 Chulalongkorn University students.

3. The domain-referenced model of item analysis was used to establish test-item sensitivity indices and facility values using the pre-post test data. Swaminathan, Hambleton and Algina's (In Pitiyanuwat, 1986) methods of test analysis were used to find the reliability and validity indices of the tests. Item-objective congruence was also carried out to established the test content validity.

12

(See Appendix III for the formulae of test analyses)

4. Test administration of the main sample groups for

the research consisting of 1,260 lower secondary, 994 upper secondary and 493 tertiary (Chulalongkorn University) students using the nine standardized and validated domain-referenced tests of English use was carried out. The data were analyzed using the Rasch Model to find item-fit statistics which indicated that over 95% of the items fit the ICC.

***Step Five***

To validate the tests the procedures employed consisted of the following:

1. As stated in # 3 of Step Four the domain-referenced methods of test analysis were used to establish the tests' reliability and validity indices.

2. The validity established through the item-objective congruence method was content validity.

3. The tests' criterion-related validty was established through the pre-post test method mentioned in #3 of Step Four. Besides the pre-post test method of test analysis, the tests'criterion-related validity was also established using t-tests for secondary students divided into 2 groups on the basis of school type--small-medium sized, and large-very large sized, and F-tests for university students divided into 3 groups on the basis of their year of study--first-year, second-year, and third-year. The t-values and the F-ratios which were statistically significant at .05 and .01 levels confirmed the tests' criterion-related validity.

4. The tests' construct validity was established using Pearson Product Moment Correlation method. Correlational matrices among the various parts of the tests using selection and supply item types and measuring different traits via unitary skill mode and integrative skill mode were analyzed. The correlations were labelled as follows:

C = Monotrait-Heteromethod

H = Heterotrait-Heteromethod

MM = Monotrait-Monomethod

V = Total score with part score

M = Heterotrait-Monomethod

Correlations labelled C should be greater than those labelled M and MM which, in turn, should be greater than correlations labelled MM if the tests are convergently valid. Besides, C and MM correlations should be greater than H if the tests are discriminantly valid (Campbell and Fiske, 1967). The V correlations should also be greater than the MM to prove the tests' discriminant validity.

The correlational matrices for the three sets of domain-referenced tests exhibited these patterns of correlations indicating the tests' construct validity of the convergent and discriminant types.

***Step Six***

Domain-content standard scores of the three tests were computed using the percentile method. Groups of items were identified from 5 to 95 percentiles. The grouping of the scores from each test exhibited similar patterns indicating the

consistency of item difficulty and student ability.

14

(See Appendix IV for Domain-Content Standard Scores)

***Step Seven***

The generalizability of the domain-referenced tables of specification was proven by F-tests of the item difficulty values. The F-ratio of 0.00 and the probability of 0.99 on all the tests proved that the ICC curves of the three parallel tests for each level did not statistically significantly deviate from the standard ICC.

(See Appendix V for the F-tests of the domain-referenced tests' ICC with the standard ICC)

***LANGUAGE TESTING RESEARCH IN THE SECOND SENSE***

The research intended to find out about the nature and specific characteristics of communicative proficiency in English by using the constructed, standardized and validated domain-referenced tests of English use. The research procedures consisted of:

***Step One***

The analyses using Pearson Product Moment Correlation method revealed that language components,ie., phonology, vocabulary andgrammar, weremoderately and significantly corrrelated with language modalities, ie., oral/aural and graphic/visual, and that the shared variances with the oral/aural or sound modality were 19%, 35% and 24% for the lower secondary, upper secondary, and tertiary levels respectively. The variances shared with the visual modality were 22%, 24%, and 15% respectively.

*Step Two*

Exploratory principal component factor analyses using



varimax rotation were performed on the data obtained through the multitrait-multimethod assessment scheme. For the lower secondary (students aged 14-15) four points emerged:

1. At this level the sound modality was finely differentiated into 3 factors: **Factor One** --listening, word-stress, and speaking (SSQ 1.458), **Factor Two**--integrative skills of listening-reading-writing and sound discrimination (SSQ 1.193) and **Factor Three**--integrative skills of listening and writing (SSQ 1.129). Thus, the shared variances of sound modality with total communicative proficiency were 11.2%, 9.0% and 8.5% totalling 28.7%. **Factor Four** was visual modality consisting of writing, reading, vocabulary, reading-writing, grammar and speaking (measured indirectly via reading).

2. Vocabulary and grammar were grouped with the visual modality indicating that the test mainly measured this factor at this level of education.

3. The visual modality factor shared common variance with the total variance of communicatively proficiency by 25%.

4. The four factors could together explain domain-referenced communicative proficiency by 53.7%

Factor analysis performed on test score data of upper-secondary students(students aged 17-18) indicated that:

1. at this level of education, communicative

proficiency in English consisted of two factors. **Factor One** was sound modality consisting of listening, word-stress, and speaking assessed through a self-rating method. **Factor Two** was visual modality consisting of vocabulary and speaking measured indirectly via reading.

16

2. while vocabulary was grouped with visual variables, grammar was grouped with sound variables. Again the indirect test of speaking via reading was grouped under the visual modality factor.

For university level (students aged 17/18-20/21) the findings indicated that:

1. domain-referenced communicative proficiency was differentiated into three groups of factors with one sound modality factor and two visual modality factors. **Factor One** of the visual modality consisted of indirect speaking measured via reading, grammar, reading, and vocabulary, **Factor Two** of the visual modality consisted of integrative skills of reading-writing and unitary skill of reading. **Factor Three** was sound modality consisting of sound discrimination, word-stress, speaking measured via self-rating, integrative skills of listening and writing, and listening-reading-writing.

2. visual modality factors combined could explain the variance of communicative proficiency by 20.7% while sound modality could explain it by 31.5%.

3. similar to the findings for the lower secondary level, grammar and vocabulary were grouped with visual modality.

4. similar to the findings for both levels of secondary education, the findings for this level confirmed the grouping of indirect speaking measured via reading with visual modality. (See Appendix VI for factor analysis results)

*Conclusions*

From the aforementioned findings for the three

17

levels of education, the following conclusions about the nature and specific characteristics of communicative proficiency could be drawn:

1. At the lower intermediate level of proficiency or the lower secondary level of education, sound modality factors were most finely differentiated. Conversely, at the upper intermediate level or tertiary level of education the visual modality factors were most finely differentiated. At the intermediate level of proficiency or upper secondary level of education each modality had equally only one factor.

The findings might be the results of the nature or process of English acquisition of Thai students as well as the emphasis differently placed on each modality. At the lower intermediate level, oral-aural modes of communication are stressed most whereas at the upper-intermediate level, visual modality especially the reading skills receive most emphasis.

2. Language component variables of sound, vocabulary and grammar were never once grouped together into a

differentiated factor. They were always subsumed under the skill factors--sound discrimination and word-stress with aural /oral skills and grammar and vocabulary sometimes together under the visual modality factor for lower secondary and tertiary levels, and sometimes separately--vocabulary with visual modality and grammar with sound modality for upper secondary level.

3. The sound factors across the three levels of education could explain the variance of domain-referenced

18

communicative proficiency by 28.7%, 36.9% and 31.5% respectively, while the visual modality factors could explain it by 25.0%, 9.1% and 20.7% respectively, or together at the ratio of 1.8:1. This illustrated the significant role of sound variables in student acquisition of communicative proficiency as they appeared to have more common variance with that of total communicative proficiency.

***LANGUAGE TESTING RESEARCH IN THE THIRD SENSE***

This research also aimed at discovering the levels of student Enlish communicative proficiency in the specified domains in the three levels of education. Parametric statistical analyses were performed to find the mean, standard deviation, kurtosis, skewness and coefficient of variation (CV) of the tests. The 5 criterion levels were used to identify of student levels of communicative proficiency--

1 = very poor, need maximal remedial work

2 = rather poor, need remedial work

3 = average, still need remedial work

4 = rather good, need some remedial work

5 = very good, do not need any remedial work

The findings indicated that for the lower secondary students their unitary skills of reading, listening, speaking were in general average while their integrated skills were quite weak. The CV indicated a high degree of score variation ( 2-7).

For the upper secondary students their speaking

19

skills were average while their other unitary and integrated skills ranged from rather poor to average. The CV indicated an average degree of score variation (2-5).

For university students, they were average to good in speaking, average in reading, rather poor to average in the integrated skills. The CV indicated an average degree of score variation (2-5).

Across the levels, Thai students ranged from levels 2-4 in their unitary skills while their integrated skills ranged from levels 2-3. The CV indicated that lower secondary students had the greatest degree of score variation (2-7) while the upper secondary and university groups' had average degree--2-5 and 1-4 respectively. (See Appendix VII for students' levels of communicative proficiency in English)

## RECOMMENDATIONS

The research on domain-referenced testing in both phases led to the following recommendations:

### *For Further Research*

1. Research into the degrees of contribution that language components have on acquisition of the unitary and integrated language skills should be conducted as well as research to find the most effective remedial and instructional processes for different levels of proficiency.

2. Amplified objectives, domain-referenced test tables of specification and domain-referenced test items should be developed, standardized and validated for every sub-level of primary, secondary, and tertiary education in Thailand.



### *For Practical Applications in English Teaching and Learning*

1. Language components--phonology, grammar, vocabulary--should not be presented or practised without the context of communication as they were never once grouped into a separate factor.

2. At the beginning to lower intermediate levels or with students aged 7 to 14 the sound modality should get more emphasis than the visual one if language acquisition is to occur naturally.

3. At the intermediate levels or with students aged 15 to 17, both modalities may receive equal amount of presentation and practice.

4. At the upper intermediate to advanced levels or

with students aged17/18 to20/21 the unitary skill or integrated skills of reading and writing should be emphasized most.

5. As the sound modality factors shared greater variance with communicative proficiency (at the ratio of 1.8:1), media and materials for listening, speaking, listening-speaking, listening-writing, and listening-reading-writing skills should be provided for both classrooom instruction and self-access learning for individual students in all levels of education.



## *REFERENCES*

Alkin, M. C. (ed.) 1973. Domain-Referenced Curricular Evaluation: A Technical Handbook and a Case Study from the Minnemast Projects. CSE, Monograph Series in Education, Center for the Study of Evaluation, UCLA.

Bachman, L. F. & Palmer, A. S. 1983. "The Construct Validity of the FSI Oral Interview." In Oller, J. W., Jr. (ed.) Issues in Language Testing Research. NJ: Newbury.

^^^^^^^^^^. 1985. "Source of Variance in Language Test Scores".

Paper presented at the National Seminar on Problems and Issues in Language Testing and Evaluation at Tertiary Level, 15 May - 4 June, 1985. Chulalongkorn University Language Institute.

^^^^^^^^^^. 1986. "The Development and Use of Criterion-Referenced Tests of Language Proficiency in Language Program Evaluation." Paper prersented at the International Seminar on Trends in Language Programme Evaluation, 9 - 11 December, 1986. Chulalongkorn University Language Institute.

Baker, E. L. 1974. "Beyond Objectives: Domain-Referenced Tests for Evaluation and Instructional Improvement." Domain-Referenced Testing. Englewood Cliffs, NJ: Educational Technology.

^^^^^^^^^^. 1978. " Is Something Better than Nothing? The Metaphysical Aspects of Test Design." In Baker, E. L. & Quellmalz, E. S. (eds.). Educational Testing and



Evaluation: Design, Analysis, and Policy. Beverly Hills, CA: Sage.

Campbell, D. T. & Fisks, D. W. 1967. "Convergent and Discriminant Validation by the Multitrait- Multimethod Matrix." In Mehrens, W. A. & Ebel, R. L. (eds.) Principles of Educational and Psychological Measurement. Chicago: Rand McNally.

Canale, M & Swain M. 1981. "A Theoretical Framework for Communicative Competence." In Palmer, A. S., Groot, P. J.

M. & Trosper, G. A. (eds.) The Construct Validation of Tests of Communicative Competence. Washington, D.C.: Teachers of English to Speakers of Other Languages.

^^^^^^^^^^. 1983. "From Communicative Competence to Communicative Language Pedagogy." In Richards, J. C. & Schmidt, W. R. (eds.) Language and Communication. London: Longman.

Carroll, B. J. 1980. Testing Communicative Performance. Oxford: Pergamon.

Cronbach, L. J. 1951. "Coefficient Alpha and the Internal Structure of Tests." Psychometrika, 16.

^^^^^^^^^^. & Meehl, P. E. 1972. "Construct Validity in Psychological Tests." In Noll, V. H. et. al. (eds.) Introductory Readings in Educational Measurement. Boston: Houghton Mifflin.

Cziko, G. A. 1983. "Psychometric and Edumetric Approaches to Language Testing." In Oller, J. W., Jr. (ed.) Issues in Language Testing Research. Rowley, MA: Newbury.

Duncan, A. D. 1974. "Tracking Behavioral Growth: Day-to-Day Measures of Frequency over Domains of Performance." In Hively, W. (ed.) Domain-Referenced Testing. Englewood Cliffs, NJ: Educational Technology Publication.

23

Fan, C. T. 1952. Item Analysis Table. Princeton, NJ: Educational Testing Service.

Farhady, H. 1983. "On the Plausibility of the Unitary Language Proficiency Factor." In Oller, J. W., Jr. (ed.) Issues in Language Testing Research. Rowley, MA: Newbury.

Higgs, T. V. & Clifford, R. 1982. "The Push toward Communication."

In Higgs, T. V. (ed.) Curriculum Competence and the Foreign Language Teacher. The American Council on the Teaching of Foreign Languages. Foreign Language Education Series. Lincolnwood, IL: National Textbook Co.

Hinofotis, F. B. 1983. "The Structure of Oral Communication in an Educational Environment: A Comparison of Factor Analytic Rational Procedures." In Oller, J. W., Jr. (ed.) Issures in Language Testing Research. Rowley, MA: Newbury.

Hively, W., Patterson H. L. & Page, S. A. 1968. "Universe Defined System of Arithmetic Achievement Test." Journal of Educational Measurements, 5.

^^^^^^^^^^. 1974. "Introduction to Domain-Referenced Testing." In Hively, W. (ed.) Domain-Referenced Testing. Englewood Cliffs, NJ: Educational Technology Publications.

^^^^^^^^^^. & Reynolds, M. C. (eds.) 1975. Domain-Referenced Testing in Special Education. The Leadership Training Institute/ Special Education.

Johnson, T. J. 1974. "Program and Product Evaluation from a Domain-Referenced Viewpoint." In Hively, W. (ed.) Domain-Referenced Testing. Englewood Cliffs, NJ: Educational Technology Publications.



Lord, F. M. & Novick, M. R. 1968. Statistical Theories of Mental Test Scores. Reading, MA: Addison & Wesley.

Mager, R. F. 1962. Preparing Instructional Objectives. Palo Alto, CA: Fearon Publishers.

McBride, J. R. & Weise, D. J. 1974. A Word Knowledge Item Pool for

Adaptive Ability Measurement. Research Report 74-2, Psychometric Methods Program, Dept. of Psychology, University of Minnesota, June.

Miller, D. M. 1974. "Content, Items, Decisions: Orienting Curriculum Assessment Survey to Curriculum Management and Modification." In Hively, W. (ed.) Domain-Referenced Testing. Englewood Cliffs, NJ: Educational Technology Publications.

Millman, J. 1974. "Sampling Plans for Domain-Referenced Tests." In Hively, W. (ed.) Domain-Referenced Testing. Englewood Cliffs, NJ: Educational Technology Publications.

Morrow, K. 1977. Techniques of Evaluation for a Notional Syllabus. Centre for Applied Language Study: University of Reading.

Nitko, A. J. 1982. Educational Tests and Measurement: An Introduction. New York: Harcourt, Brace, Jovanovich.

Oller, J. W., Jr. 1969. Extract from An Interview with John Oller on Language Testing Today. (Video Interview with Ann Newton). USIS, AUA Library, Bangkok.

**********. 1979. Language Tests At Schools: A Pragmatic Approach. London: Longman.

**********. (ed.) 1983. Issues in Language Testing Research. Rowley, MA: Newbury.

Osburn, H. G. 1968. "Item Sampling for Achievement Testing." Educational and Psychological Measurement, 28, 95-104.

25

**********. 1969. "The Effect of Item Stratification on Errors of Measurement." Educational and Psychological Measurement, 29, 295-31.

Palmer, A. S. Groot, P. J. M. & Trosper, G. A. (eds.) 1981. The Construct Validation of Tests of Communicative Competence. Washington, D. C.: Teachers of English to Speakers of Other Languages.

Pitiyanuwat, S. 1986. "Domain-Referenced Testing." Selected Papers on Testing. Center for National Educational Development for Thailand, Ministry of University Affairs. Nonthaburi: Sukhothaithammathirat Open University Press. (In Thai).

Popham, W. J. 1978. Criterion-Referenced Measurement. Englewood Cliffs, NJ: Prentice Hall.

^^^^^^^^^^. 1981. Modern Educational Measurement. Englewood Cliffs, NJ.: Prentice Hall.

Purves, A. C. et. al. 1984. "Towards a Domain-Referenced System for Classifying Composition Assignments." Research in the Teaching of English. 18 (4), December, NCTE.

Rabehl, G. J. 1971. Methodological Characteristics of Instructional Accountability at Osseo. Unpublished paper available from Osseo, minnesota Public Schools. (Referred to in Sension, D. B. & Rabehl, G. J., 1974).

Roid, G. & Haladyna, T. 1982. A Technology for Tests Item Writing. New York: Academic Press.

Scannell, D. P. & Tracy, D. B. 1975. Testing and Measurement in the Classroom. Boston: Houghton Mifflin.

Sension, D. B. & Rabehl, G. J. 1974. "Test-Item Domains and Instructional Accountability." In Hively, W. (ed.)



Domain-Referenced Testing. Englewood Cliffs, NJ: Educational

Technology Publications.

Slaser, R. 1978. Advances in Instructional Psychology. New York: Halsted Press.

Thorndike, R. L. 1982. Applied Psychometrics. Boston: Houghton Mifflin.

Valette, R. M. & Disick, R. S. 1972. Modern Language Performance Objectives and Individualization: A Handbook. New York: Harcourt-Brace Jovanovich.

^^^^^^^^^^. 1977. Modern Language Testing. ( 2nd ed.) New York: Harcourt-Brace Jovanovich.

Vollmer, H. J. & Sang, F. 1983. "Competing Hypothesis about Second Language Ability: A Plea for Caution." In Oller, J. W., Jr.(ed.) Issues in Language Testing Research. Rowley, MA: Newbury.

Warm, T. A. 1978. A Primer for Item Research Theory. Dept. of Transportation, Coast Guard, Technical Report 9412 78. Oklahoma City: US Coast Guard Institute.

Whitney, D. R. & Sabers, D. L. 1970. "Improving Essay Examinations III. Use of Item Analysis." Technical Bulletin, 11. Iowa City: University Evaluation and Examination Service. (mimeo.)

Wibulsri, Y. 1983. Fundamentals of Achievement Test Construction. Bangkok: Dept. of Educational Research, Faculty of Education, Chulalongkorn University. (In Thai).

Widdowson, H. G. 1972. "The Teaching of English as Communication." English Language Teaching, 27 (1), October, 1972.

^^^^^^^^^^. 1978. Teaching Language as Communication. London:

Oxford University Press.

Wright, B. D. & Stone, M. H. 1979. *Best Test Design.* Chicago: Mesa Press.

Yalden, J. 1983. *The Communicative Syllabus: Evaluation, Design and Implementation.* Oxford: Pergamon Press.

## *APPENDIX I*

### *Amplified Objectives for English Core Course: E 615, 616 (Upper Secondary Level)*

1. After listening to the economic, political, and sports news, the student is able to answer four-choice multiple choice question about the main ideas and details of the news.
2. After listening to the news, the student is able to make an oral summary of the news.
3. While listening to descriptions of people, animals, objects and places, the student is able to ask for more information about the objects of the description.
4. While listening to narrations about incidents such as an accident, the student is able to take notes of what happened and able to sequence the incidents.
5. After listening to songs and poems, the student is able to express his/her evaluative comments on what he/she has heard.
6. When someone makes a wrong statement about something, e.g., a personal relationship, the student is able to correct it.
7. The student is able to ask for details about people, objects, places, time and activities, and is able to answer Wh questions accordingly.
8. The student is able to make an appropriate oral invitation to various groups of people on various occasions.

9. The student is able to speak about his determination to do something in an appropriate context or situation, e.g., when asked about his/her future plan for study.

10. When a classmate speaks about something in which the student is not interested, he/she is able to express his disinterest.

11. After reading a letter or a note giving directions to a particular place, the student is able to draw a map of the place.

12. After reading about a person's daily routine, the student is able to complete a time-table about that person's activities.

13. After reading captions, the student is able to match the captions with the appropriate pictures.

14. After reading an assigned short story or novel, the student is able to relate it to others.

15. After reading a congratulatory letter or card, the student is able to write a thank-you reply.

16. After listening to a telephone message, the student is able to take down the message.

17. After reading a short story, an article or a novel, the student is able to summarize it.

18. After reading a short story or a novel, the student is able to extend it from his imagination.

19. After reading a "Want" ad, the student is able to write a letter of application.

20. After receiving good or bad news, the student is able to write a letter of congratulation or sympathy to the person concerned.

## *APPENDIX II*

### *Specimen Domain-Referenced Test Table of Specification*

| Obj. | Language Function | Mode | Genre | Channel | Style | Role | Status | Topic |
|---|---|---|---|---|---|---|---|---|
| 1 | Getting information | List-Speak | News | Radio/ TV | Formal | Stu. to Peers & Teachers | Equal | Econ. Polit. Sports |
| 5 | Express Opinion | Speak/ Writ. | Literary | Face to Face/ Radio | Formal | Student to Peers &Teacher | Equal | Point of view |
| 17 | Summarize | Read/ Writ. | Short story &Novel | Book/ Magazine | Formal | Student to Peers &Teacher | Equal | Adventure |

**Test Item Type and Test Technique**

Item 1 Selection type--Four-choice Multiple Choice

Length of News--50-100 words

Item 5 Supply type--Integrative skills--Listening/Speaking and Listening/Writing

Length of Input(song and poem) 150-200 words

Length of Output(Student expression) 100-150 words

Item 17 Supply type--Integrative skills--Reading/Writing

Length of Input (Short story and article ) 1000-2000 words

(Novel) 10000-15000 words

Length of Output (Student summaries) For short story novel--500-600 words

For novel--1000-1500 words

Marking Scheme

Item 1 1 for correct

o for incorrect

Item 5 On a Rating Scale of 7 from 1=no competence, 2=minimal competence, 3=threshold competence

4=some communicative competence

5=quite competent

6=good-very good communicative competence

7=near educated native communicative competence

Item 17 Same as Item 5

## *APPENDIX III*

### *Formulae for Item Analyses*

1. Whitney & Sabers Multiple-point Item Analysis

Facility Value (FV) = $S_u + S_l - (N_T) X_{min}$

$$N_T(X_{max} - x_{min})$$

$S_u$ = total scores of the upper group

$S_l$ = total scores of the lower group

$N_T$ = total number of testees used in the analysis

$X_{max}$ = highest score obtained

$X_{min}$ = lowest score obtained

Discrimination Index (D)

$D = S_u - S_l$

^^^^^^^^^^^^^^^^^

$$N_u (X_{max} - X_{min})$$

$S_u, S_l, X_{max}, X_{min}$ symbolize the same values as in the formula for FV

$N_u$ = number of testees in the upper group

**Source:** Whitney & Sabers, 1970, 1971 In Scannell & Tracy, 1975: 221-224.

2. Domain-Referenced Test Item Analysis

Facility Value (FV) = Proportion of testees passing the each item after instruction

$$\text{Sensitivity Index}(S_i) = \frac{R\ post_i - R\ pre_i}{T}$$

$R\ post_i$ = Number of testees passing $item_i$ after instruction

$R\ pre_i$ = Number of testees passing the item ($item_i$) before instruction

T = Total number of testees

The values of $S_i$ range between -1 and +1. Good items should have the values close to +1.

**Source:** Pitiyanuwat, 1986: 79-80.

3. Index of Item-Objective Congruence (Content Validity)

$$I_{io} = \frac{(k - 1)s_o - S_{oi}}{2N (k - 1)}$$

$I_{io}$ = Index of item objective congruence for item$_i$ and Objective$_o$

K = Number of objectives to be measured by the test

N = Number of specialist raters

$S_o$ = Total scores of all raters for objective $_o$

$S_{io}$ = Total scores of all raters for all objectives except objective $_o$

**Source:** Wibulsri, 1983.

4. Swaminathan, Hambleton & Algina Coefficient of Reliability

$$K = \frac{P_o - P_e}{1 - P_e}$$

K = Coefficient of Reliability

$P_o$ = Observed Proportion (Proportion of testees passing both pre and post tests + proportion of testees failing both pre and post tests)

$P_e$ = Expected Proportion [ (Proportion of testees passing both tests + Proportion of testees passing the pre-test but failing the post test) x (Proportion of

testees passing both tests + Proportion of testees failing the pre test but passing the post test) +

(Proportion of testees failing the pre test but passing the post test)

---

(Proportion of testees failing both tests) x (Proportion of testees passing the pre test but failing the post test + Proportion of testees failing both tests)]

Example: Criterion of Passing = 80%

| Pre / Post | Pass | Fail | Total Pass |
|---|---|---|---|
| Pass | .66 | .14 | .80 |
| Fail | .04 | .16 | .20 |
| Total | .70 | .30 | 1.00 |

$P_o$ = .66 + .16 = .82

$P_e$ = (.70)(.80) + (.30) + (.20) = .62

$$K = \frac{.82-.62}{1-.62}$$

$$= .53$$

Domain-Referenced Validity Coefficient

$$V = \frac{a + c}{}$$

$$V = \frac{a + b}{a + b + c + d}$$

V = Validity Coefficient

a = Number of people passing the criterion after instruction

b = Number of people passing the criterion before instruction

c = Number of people not passing the criterion before instruction

d = Number of people not passing the criterion after instruction

Validity coefficient ranges from 0 - 1. A value close to 1 indicates test validity.

**Source:** Pitiyanuwat, 1986: 79-81

## *APPENDIX IV*

### *The Test Domain-Content Stadard Scores*

Symbols used in the table

S = Speaking

L = Listening

R = Reading

ES=Essay or paragraph writing

DI=Dictation (Listening-Writing)

RW= Reading-Writing

LW=Listening-Writing

LR=Listening-Reading-Writing (Partial Dictation)

L = Lower Level of Secondary Education

U = Upper Level of Secondary Education

The Table reveals that for lower secondary students, the items grouped into high, middle and low percentiles measured the same skills by 40%, 65%, and 67 % respectively for Forms A, B and C averaging 57% of the items measuring the same skills for

the three forms.

For the upper level of secondary education, the test items in the three forms measured the same skills by 49% on the average.

Students in both levels of secondary education could do items measuring reading and reading best, second were items measuring their speaking skills via a self-rating method. The items they could do least were items measuring their integrative skills of listening-writing, listening-reading-wring, reading-writing and essay/paragraph writing.

***APPENDIX V***

***F-tests of the Domain-Referenced Tests' ICC (Item Characteristic***

*Curve) with the Standard ICC*

*Lower Secondary Level*

| Form | X | SD | N |
|---|---|---|---|
| A | 0.02 | 1.49 | 42 |
| B | 0.04 | 1.24 | 42 |
| C | 0.04 | 1.30 | 39 |

| | SS | DF | MS | F | P |
|---|---|---|---|---|---|
| Bet. Groups | .10 | 2 | .00 | .00 | 0.99 |
| Within Group | 211.94 | 120 | 1.81 | | |
| Total | 211.95 | 122 | | | |

*Upper Secondary Level*

| Form | X | SD | N |
|---|---|---|---|
| A | 0.06 | 0.91 | 35 |
| B | -0.02 | 1.28 | 36 |
| C | 0.10 | 1.05 | 36 |

| | SS | DF | MS | F | P |
|---|---|---|---|---|---|
| Bet. Groups | .23 | 2 | 0.11 | 0.10 | 0.99 |
| Within Group | 117.44 | 104 | 1.20 | | |
| Total | 117.67 | 106 | | | |

## *APPENDIX VI*

### *Factor Analysis Results*

Abbreviations FV = Factor for Visual Modalities--Reading, Writing, Reading-Writing, Writing, Indirect Speaking Measured via Reading

FS = Factor for Sound Modalities--Listening, Listening-Reading-Writing, Listening-Writing

SD = Sound Discrimination, ST=Word Stress

VOC = Knowledge of vocabulary, G = Use of Grammar

IS(R) = Indirect Speaking Measured via Reading

| *Lower Secondary* (N= 679) | | | | *Upper Secondary* (N = 525) | | *Tertiary* (N=493) | | |
|---|---|---|---|---|---|---|---|---|
| FV I | FS I | FS II | FS III | FS I | FV I | FSI | FVI | FVII |
| W | L | L-R-W | L-W | L | Voc | SD | IS(R) | R-W |
| R | ST | SD | | ST | R | ST | G | W |
| VOC | S | | | G | IS(R) | S | R | |
| R-W | | | | S | R-W | L-W | VOC | |
| IS(R) | | | | L-W | W | L=R-W | | |
| G | | | | L-R-W | | | | |
| SSQ | | | | SSQ | | SSQ | | |
| 2.54 | 1.458 | 1.193 | 1.128 | 3.256 | 2.447 | 2.339 | 1.974 | 1.426 |

## *APPENDIX VII*

### *Students' Levels of English Proficiency*

| *English Proficiency* | *Lower Secondary* | | *Upper Secondary* | | *Tertiary* | |
|---|---|---|---|---|---|---|
| | Level | CV | Level | CV | Level | CV |
| Language Components | 3 | 1 | 2-3 | 2 | 3 | 1 |
| Sound Modalities | 2-3 | 1 | 3 | 2 | 3 | 2 |
| Visual Modalities | 2-3 | 2 | 2-3 | 2 | 2-3 | 2 |
| Components&Modalities Combined | 2-3 | 1-7 | 2-3 | 2-5 | 3 | 1-4 |